SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT:
       (A) NAME: Hoechst Aktiengesellschaft
       (B) STREET: -
       (C) CITY: Frankfurt
       (D) STATE: -
       (E) COUNTRY: Germany
       (F) POSTAL CODE (ZIP): 65926
       (G) TELEPHONE: 069-305-7072
       (H) TELEFAX: 069-35-7175
       (I) TELEX: -

    (ii) TITLE OF INVENTION: Purification of higher order transcription
       complexes from transgenic non-human animals

    (iii) NUMBER OF SEQUENCES: 17

    (iv) COMPUTER READABLE FORM:
       (A) MEDIUM TYPE: Floppy disk
       (B) COMPUTER: IBM PC compatible
       (C) OPERATING SYSTEM: PC-DOS/MS-DOS
       (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 12 amino acids
       (B) TYPE: amino acid
       (C) STRANDEDNESS: single
       (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (ix) FEATURE:
       (A) NAME/KEY: Peptide
       (B) LOCATION: 1..12

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Val
1   5   10

(2) INFORMATION FOR SEQ ID NO: 2:

    (i) SEQUENCE CHARACTERISTICS:
       (A) LENGTH: 11 amino acids

              (B) TYPE: amino acid
              (C) STRANDEDNESS: single
              (D) TOPOLOGY: linear

 5   (ii) MOLECULE TYPE: peptide

          (ix) FEATURE:
              (A) NAME/KEY: Peptide
10   (B) LOCATION: 1..11

          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

15   Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
          1   5   10

     (2) INFORMATION FOR SEQ ID NO: 3:
20
          (i) SEQUENCE CHARACTERISTICS:
              (A) LENGTH: 10 amino acids
              (B) TYPE: amino acid
              (C) STRANDEDNESS: single
25   (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: peptide

30   (ix) FEATURE:
              (A) NAME/KEY: Peptide
              (B) LOCATION: 1..10

35   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

          Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
          1   5   10

40
     (2) INFORMATION FOR SEQ ID NO: 4:

          (i) SEQUENCE CHARACTERISTICS:
              (A) LENGTH: 9 amino acids
45   (B) TYPE: amino acid
              (C) STRANDEDNESS: single
              (D) TOPOLOGY: linear

          (ii) MOLECULE TYPE: peptide
50

          (ix) FEATURE:
              (A) NAME/KEY: Peptide
              (B) LOCATION: 1..9
55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1   5

(2) INFORMATION FOR SEQ ID NO: 5:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 22 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: cDNA

 (ix) FEATURE:
  (A) NAME/KEY: exon
  (B) LOCATION: 1..22

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGAGCAACCG CCTGCTGGGT GC   22

(2) INFORMATION FOR SEQ ID NO: 6:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 21 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

 (ii) MOLECULE TYPE: cDNA

 (ix) FEATURE:
  (A) NAME/KEY: exon
  (B) LOCATION: 1..21

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCTGTGTTGC CTGCTGGGAC G   21

(2) INFORMATION FOR SEQ ID NO: 7:

 (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 21 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

44

```
       (ii) MOLECULE TYPE: cDNA


       (ix) FEATURE:
   5          (A) NAME/KEY: exon
              (B) LOCATION: 1..21


       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
  10
   GGAGACTGAA GTTAGGCCAG C                                        21


   (2) INFORMATION FOR SEQ ID NO: 8:
  15
       (i) SEQUENCE CHARACTERISTICS:
              (A) LENGTH: 76 base pairs
              (B) TYPE: nucleic acid
              (C) STRANDEDNESS: single
  20          (D) TOPOLOGY: linear

       (ii) MOLECULE TYPE: cDNA


  25   (ix) FEATURE:
              (A) NAME/KEY: exon
              (B) LOCATION: 1..76


  30   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

   GCGGCACCAG GCCGCTGCTG TGATGATGAT GATGATGGCT GCTGCCCATG ACTGCGTAAT    60

   GCGGTCATGA CGCTTT                                                76
  35

   (2) INFORMATION FOR SEQ ID NO: 9:

       (i) SEQUENCE CHARACTERISTICS:
  40          (A) LENGTH: 75 base pairs
              (B) TYPE: nucleic acid
              (C) STRANDEDNESS: single
              (D) TOPOLOGY: linear

  45   (ii) MOLECULE TYPE: cDNA


       (ix) FEATURE:
              (A) NAME/KEY: exon
  50          (B) LOCATION: 1..75


       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

  55   GAAGGGGGTG GGGGAGGCAA GGGTACATGA GAGCCATTAC GTCGTCTTCC TGAATCCCTT    60
```

45

TAGCCGCTTT GCTCG   75

(2) INFORMATION FOR SEQ ID NO: 10:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 22 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA

        (ix) FEATURE:
            (A) NAME/KEY: exon
            (B) LOCATION: 1..22

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCCTATGACG TCCCGGATTA CG   22

(2) INFORMATION FOR SEQ ID NO: 11:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 22 base pairs
            (B) TYPE: nucleic acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: cDNA

        (ix) FEATURE:
            (A) NAME/KEY: exon
            (B) LOCATION: 1..22

        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTGGAGTGGT GCCCGGCAAG GG   22

(2) INFORMATION FOR SEQ ID NO: 12:

        (i) SEQUENCE CHARACTERISTICS:
            (A) LENGTH: 19 amino acids
            (B) TYPE: amino acid
            (C) STRANDEDNESS: single
            (D) TOPOLOGY: linear

        (ii) MOLECULE TYPE: peptide

```
(ix)  FEATURE:
       (A)  NAME/KEY: Peptide
       (B)  LOCATION: 1..19


(xi)  SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
1               5                       10                  15

Arg Gly Cys
```

(2) INFORMATION FOR SEQ ID NO: 13:

```
       (i)   SEQUENCE CHARACTERISTICS:
              (A)  LENGTH: 1310 base pairs
              (B)  TYPE: nucleic acid
              (C)  STRANDEDNESS: single
              (D)  TOPOLOGY: linear

       (ii)  MOLECULE TYPE: cDNA


       (ix)  FEATURE:
              (A)  NAME/KEY: exon
              (B)  LOCATION: 1..1310


       (xi)  SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CCATGGGCTA TCCCTATGAC GTCCCGGATT ACGCAGTCAT GGGCAGCAGC CATCATCATC    60

ATCATCACAG CAGCGGCCTG GTGCCGCGCG GCAGCCATAT GGATCAGAAC AACAGCCTGC   120

CACCTTACGC TCAGGGCTTG GCCTCCCCTC AGGGTGCCAT GACTCCCGGA ATCCCTATCT   180

TTAGTCCAAT GATGCCTTAT GGCACTGGAC TGACCCCACA GCCTATTCAG AACACCAATA   240

GTCTGTCTAT TTTGGAAGAG CAACAAAGGC AGCAGCAGCA ACAACAACAG CAGCAGCAGC   300

AGCAGCAGCA GCAGCAACAG CAACAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC   360

AGCAGCAGCA GCAGCAGCAA CAGGCAGTGG CAGCTGCAGC CGTTCAGCAG TCAACGTCCC   420

AGCAGGCAAC ACAGGGAACC TCAGGCCAGG CACCACAGCT CTTCCACTCA CAGACTCTCA   480

CAACTGCACC CTTGCCGGGC ACCACTCCAC TGTATCCCTC CCCCATGACT CCCATGACCC   540

CCATCACTCC TGCCACGCCA GCTTCGGAGA GTTCTGGGAT TGTACCGCAG CTGCAAAATA   600

TTGTATCCAC AGTGAATCTT GGTTGTAAAC TTGACCTAAA GACCATTGCA CTTCGTGCCC   660

GAAACGCCGA ATATAATCCC AAGCGGTTTG CTGCGGTAAT CATGAGGATA AGAGAGCCAC   720
```

47

```
GAACCACGGC ACTGATTTTC AGTTCTGGGA AAATGGTGTG CACAGGAGCC AAGAGTGAAG        780

AACAGTCCAG ACTGGCAGCA AGAAAATATG CTAGAGTTGT ACAGAAGTTG GGTTTTCCAG        840

CTAAGTTCTT GGACTTCAAG ATTCAGAACA TGGTGGGGAG CTGTGATGTG AAGTTTCCTA        900

TAAGGTTAGA AGGCCTTGTG CTCACCCACC AACAATTTAG TAGTTATGAG CCAGAGTTAT        960

TTCCTGGTTT AATCTACAGA ATGATCAAAC CCAGAATTGT TCTCCTTATT TTTGTTTCTG       1020

GAAAAGTTGT ATTAACAGGT GCTAAAGTCA GAGCAGAAAT TTATGAAGCA TTTGAAAACA       1080

TCTACCCTAT TCTAAAGGGA TTCAGGAAGA CGACGTAATG GCTCTCATGT ACCCTTGCCT       1140

CCCCCACCCC CTTCTTTTTT TTTTTTTAAA CAAATCAGTT TGTTTTGGTA CCTTTAAATG       1200

GTGGTGTTGT GAGAAGATGG ATGTTGAGTT GCAGGGTGTG GCACCAGGTG ATGCCCTTCT       1260

GTAAGTGCCC CTTCCGGCAT CCCGGAATTC CTGCAGCCCA ACGCGGCCGC                  1310
```

(2) INFORMATION FOR SEQ ID NO: 14:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4286 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION: 1..4286

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
GAATTCCCCT GCAGGTCACT TAGCGTTGGC CACATAGTAG GTTCTCAAAT ACTTGTTAAT         60

AAATAAGTTT GTTCGAGAAG CTGGGCAATG ATATTCTACA GCTGGAAGAA GAAACATAAT        120

GATCTAGTAA TTAGCTCAAT TAAAAATAAA CGTTCTTCTT TCCTCAGAGG AGCATTTCCC        180

AAGGCCTGCC TTGATAGCCA TCCAAAAAGG CCAAGCTCAT CCAATCTTGC CCTAGATTTA        240

TGCTAAAATG CAGTTACAAT CGATAGGATG ACAGAAAACG ACAGCACTTA TTTAAATATA        300

ATAGGCACTT ATTTAAATAG GAGAAGCTGT GACTTCATAG CAAGTGTTGG GGTTAGGAAA        360

CTGGGTGGAT AAACTTGCTG ATGCTGTAGA TCTTAGCCTC TACATGAGAT CATGTGGAAA        420

ATCTGAAAGC ATTTTAGGTT CCTTATGTTT GCAATCAAAT AACTGTACAC CTTTTAATTT        480

AAAAAGTACC ATGAGGCACA CACACACACT CGCAGGAACT TTTTGGCGTA ACAAAACTAG        540
```

48

```
AATTAGATCT AAAAGCTAAC TGTAGGACTG AGTCTATTCT AAACTGAAAG CCTGGACATC      600

TGGAGTACCA GGGGGAGATG ACGTGTTACG GGCTTCCATA AAAGCAGCTG GCTTTGAATG      660

GAAGGAGCCA AGAGGCCAGC ACAGGAGCGG ATTCGTCGCT TTCACGGCCA TCGAGCCGAA      720

CCTCTCGCAA GTCCGTGAGC CGTTAAGGAG GCCCCCAGTC CCGACCCTTC GCCCCAAGCC      780

CCTCGGGGTC CCCGGGCCTG GTACTCCTTG CCACACGGGA GGGGCGCGGA AGCCGGGGCG      840

GAGGAGGAGC CAACCCCGGG CTGGGCTGAG ACCCGCAGAG GAAGACGCTC TAGGGATTTG      900

TCCCGGACTA GCGAGATGGC AAGGCTGAGG ACGGGAGGCT GATTGAGAGG CGAAGGTACA      960

CCCTAATCTC AATACAACCT TTGGAGCTAA GCCAGCAATG GTAGAGGGAA GATTCTGCAC     1020

GTCCCTTCCA GGCGGCCTCC CCGTCACCAC CCCCCCCAAC CCGCCCCGAC CGGAGCTGAG     1080

AGTAATTCAT ACAAAAGGAC TCGCCCCTGC CTTGGGGAAT CCCAGGGACC GTCGTTAAAC     1140

TCCCACTAAC GTAGAACCCA GAGATCGCTG CGTTCCCGCC CCCTCACCCG CCCGCTCTCG     1200

TCATCACTGA GGTGGAGAAG AGCATGCGTG AGGCTCCGGT GCCCGTCAGT GGGCAGAGCG     1260

CACATCGCCC ACAGTCCCCG AGAAGTTGGG GGGAGGGGTC GGCAATTGAA CCGGTGCCTA     1320

GAGAAGGTGG CGCGGGGTAA ACTGGGAAAG TGATGTCGTG TACTGGCTCC GCCTTTTTCC     1380

CGAGGGTGGG GGAGAACCGT ATATAAGTGC AGTAGTCGCC GTGAACGTTC TTTTTCGCAA     1440

CGGGTTTGCC GCCAGAACAC AGGTAAGTGC CGTGTGTGGT TCCCGCGGGC CTGGCCTCTT     1500

TACGGGTTAT GGCCCTTGCG TGCCTTGAAT TACTTCCACG CCCCTGGCTG CAGTACGTGA     1560

TTCTTGATCC CGAGCTTCGG GTTGGAAGTG GGTGGGAGAG TTCGAGGCCT TGCGCTTAAG     1620

GAGCCCCTTC GCCTCGTGCT TGAGTTGAGG CCTGGCCTGG GCGCTGGGGC CGCCGCGTGC     1680

GAATCTGGTG GCACCTTCGC GCCTGTCTCG CTGCTTTCGA TAAGTCTCTA GCCATTTAAA     1740

ATTTTTGATG ACCTGCTGCG ACGCTTTTTT TCTGGCAAGA TAGTCTTGTA AATGCGGGCC     1800

AAGATCTGCA CACTGGTATT TCGGTTTTTG GGGCCGCGGG CGGCGACGGG GCCCGTGCGT     1860

CCCAGCGCAC ATGTTCGGCG AGGCGGGGCC TGCGAGCGCG GCCACCGAGA ATCGGACGGG     1920

GGTAGTCTCA AGCTGGCCGG CCTGCTCTGG TGCCTGGCCT CGCGCCGCCG TGTATCGCCC     1980

CGCCCTGGGC GGCAAGGCTG GCCCGGTCGG CACCAGTTGC GTGAGCGGAA AGATGGCCGC     2040

TTCCCGGCCC TGCTGCAGGG AGCTCAAAAT GGAGGACGCG GCGCTCGGGA GAGCGGGCGG     2100

GTGAGTCACC CACACAAAGG AAAAGGGCCT TTCCGTCCTC AGCCGTCGCT TCATGTGACT     2160

CCACGGAGTA CCGGGCGCCG TCCAGGCACC TCGATTAGTT CTCGAGCTTT TGGAGTACGT     2220
```

49

CGTCTTTAGG TTGGGGGGAG GGGTTTTATG CGATGGAGTT TCCCCACACT GAGTGGGTGG 2280

AGACTGAAGT TAGGCCAGCT TGGCACTTGA TGTAATTCTC CTTGGAATTT GCCCTTTTTG 2340

AGTTTGGATC TTGGTTCATT CTCAAGCCTC AGACAGTGGT TCAAAGTTTT TTTCTTCCAT 2400

TTCAGGTGTC GTGAGGAATT GCCCGGGGGA TCCATGGGCT ATCCCTATGA CGTCCCGGAT 2460

TACGCAGTCA TGGGCAGCAG CCATCATCAT CATCATCACA GCAGCGGCCT GGTGCCGCGC 2520

GGCAGCCATA TGGATCAGAA CAACAGCCTG CCACCTTACG CTCAGGGCTT GGCCTCCCCT 2580

CAGGGTGCCA TGACTCCCGG AATCCCTATC TTTAGTCCAA TGATGCCTTA TGGCACTGGA 2640

CTGACCCCAC AGCCTATTCA GAACACCAAT AGTCTGTCTA TTTTGGAAGA GCAACAAAGG 2700

CAGCAGCAGC AACAACAACA GCAGCAGCAG CAGCAGCAGC AGCAGCAACA GCAACAGCAG 2760

CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGGCAGTG 2820

GCAGCTGCAG CCGTTCAGCA GTCAACGTCC CAGCAGGCAA CACAGGGAAC CTCAGGCCAG 2880

GCACCACAGC TCTTCCACTC ACAGACTCTC ACAACTGCAC CCTTGCCGGG CACCACTCCA 2940

CTGTATCCCT CCCCCATGAC TCCCATGACC CCCATCACTC CTGCCACGCC AGCTTCGGAG 3000

AGTTCTGGGA TTGTACCGCA GCTGCAAAAT ATTGTATCCA CAGTGAATCT TGGTTGTAAA 3060

CTTGACCTAA AGACCATTGC ACTTCGTGCC CGAAACGCCG AATATAATCC CAAGCGGTTT 3120

GCTGCGGTAA TCATGAGGAT AAGAGAGCCA CGAACCACGG CACTGATTTT CAGTTCTGGG 3180

AAAATGGTGT GCACAGGAGC CAAGAGTGAA GAACAGTCCA GACTGGCAGC AAGAAAATAT 3240

GCTAGAGTTG TACAGAAGTT GGGTTTTCCA GCTAAGTTCT TGGACTTCAA GATTCAGAAC 3300

ATGGTGGGGA GCTGTGATGT GAAGTTTCCT ATAAGGTTAG AAGGCCTTGT GCTCACCCAC 3360

CAACAATTTA GTAGTTATGA GCCAGAGTTA TTTCCTGGTT TAATCTACAG AATGATCAAA 3420

CCCAGAATTG TTCTCCTTAT TTTTGTTTCT GGAAAAGTTG TATTAACAGG TGCTAAAGTC 3480

AGAGCAGAAA TTTATGAAGC ATTTGAAAAC ATCTACCCTA TTCTAAAGGG ATTCAGGAAG 3540

ACGACGTAAT GGCTCTCATG TACCCTTGCC TCCCCACCC CCTTCTTTTT TTTTTTTTAA 3600

ACAAATCAGT TTGTTTTGGT ACCTTTAAAT GGTGGTGTTG TGAGAAGATG GATGTTGAGT 3660

TGCAGGGTGT GGCACCAGGT GATGCCCTTC TGTAAGTGCC CCTTCCGGCA TCCCGGATAT 3720

CCTGCAGCCC AACACGGCCG CTCGAGCATG CATCTAGAGA ACGTCACGGC CGCGATCCCC 3780

CTGTGCCTTC TAGTTGCCAG CCATCTGGTT GTTTGCCCCT CCCCCGTGCC TTCCTTGACC 3840

CTGGAAGGTG CCACTCCCAC TGTCCTTTCC TAATAAAATG AGGAAATTGC ATCGCATTGT 3900

50

```
CTGAGTAGGT GTCATTCTAT TCTGGGGGGT GGGGTGGGGC AGGACAGCAA GGGGGAGGAT     3960

TGGGAAGACA ATAGCAGGCA TGCTGGGGAT GCGGTGGGCT CTATGGGTAC CCAGGTGCTG     4020

AAGAATTGAC CCGGTTCCTC CTGGGCCAGA AAGAAGCAGG CACATCCCCT TCTCTGTGAC     4080

ACACCCTGTC CACGCCCCTG GTTCTTAGTT CCAGCCCCAC TCATAGGACA CTCAACTTGG     4140

AGCGGTCTCT CCCTCCCTCA TCAGCCCACC AAACCAAACC TAGCCTCCAA GAGTGGGAAG     4200

AAATTAAAGC AAGAAGGCTA TTAAGTGCAG AGGGAGAGAA AATGCCTCCA ACATGTGAGG     4260

AAGTAATGAT AGAAATCATA GAATTC                                         4286
```

(2) INFORMATION FOR SEQ ID NO: 15:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 3263 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: cDNA

    (ix) FEATURE:
        (A) NAME/KEY: exon
        (B) LOCATION: 1..3263

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
ATCGATAAGC TGAGATCCGG CTAGAAACTG CTGAGGGCTG GACCGCATCT GGGGACCATC      60

TGTTCTTGGC CCTGAGCGGG GCAGGAACTG CTTACCGCAG ATATCCTGTT TGCCCCAATT     120

CAGCTGTTCC ATCTGTTCTT GGCCCTGAGC GGGGCAGGAA CTGCTTACCA CAGATATCCT     180

GTTTGGCCCA TATTCAGCTG TCTCTCTGTT CCTGACCTTG ATCTGAACTT CTCTATTCTC     240

AGTTATGTAT TTTTCCCATG CCTTGCAAAA TGGCGTTACT TAAGCTAGCT TGCCAAACCT     300

ACGGCTGGGG TCTTTCACGT TTATATCTAT GAGGGGAAGG ACCCAGAGTG GGGAAGCTGG     360

GATCTTGGGA ACACGCTTCT CTACATGGCA TTGTCTGCAC GGTGGAGTCC GGATCTGAGC     420

TTGGCTTGGT TTTTAAAACC AGCCTGGAGT AGAGCAGATG GGTTAAGGTG AGTGACCCCT     480

CAGCCCTGGA CATTCTTAGA TGAGCCCCCT CAGGAGTAGA GAATAATGTT GAGATGAGTT     540

CTGTTGGCTA AAATAATCAA GGCTAGTCTT TATAAAACTG TCTCCTCTTC TCCTAGCTTC     600

GATCCAGAGA GAGACCTGGG CGGAGCTGGT CGCTGCTCAG GAACTCCAGG AAAGGAGAAG     660
```

51

```
   CTGAGGTTAC CACGCTGCGA ATGGGTTTAC GGAGATAGCT GGCTTTCCGG GGTGAGTTCT    720

   CGTAAACTCC AGAGCAGCGA TAGGCCGTAA TATCGGGGAA AGCACTATAG GGACATGATG    780

5  TTCCACACGT CACATGGGTC GTCCTATCCG AGCCAGTCGT GCCAAAGGGG CGGTCCCGCT    840

   GTGCACACTG GCGCTCCAGG GAGCTCTGCA CTCCGCCCGA AAAGTGCGCT CGGCTCTGCC    900

   AGGACGCGGG GCGCGTGACT ATGCGTGGGC TGGAGCAACC GCCTGCTGGG TGCAAACCCT    960

10 TTGCGCCCGG ACTCGTCCAA CGACTATAAA GAGGGCAGGC TGTCCTCTAA GCGTCACCAC   1020

   GACTTCAACG TCCTGAGTAC CTTCTCCTCA CTTACTCCGT AGCTCCAGCT TCACCAGATC   1080

15 CTCGAGAACG TCTCCCATGG GCTATCCCTA TGACGTCCCG GATTACGCAG TCATGGGCAG   1140

   CAGCCATCAT CATCATCATC ACAGCAGCGG CCTGGTGCCG CGCGGCAGCC ATATGGATCA   1200

   GAACAACAGC CTGCCACCTT ACGCTCAGGG CTTGGCCTCC CCTCAGGGTG CCATGACTCC   1260

20 CGGAATCCCT ATCTTTAGTC CAATGATGCC TTATGGCACT GGACTGACCC CACAGCCTAT   1320

   TCAGAACACC AATAGTCTGT CTATTTTGGA AGAGCAACAA AGGCAGCAGC AGCAACAACA   1380

25 ACAGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGCAACAG CAGCAGCAGC AGCAGCAGCA   1440

   GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA GCAACAGGCA GTGGCAGCTG CAGCCGTTCA   1500

   GCAGTCAACG TCCCAGCAGG CAACACAGGG AACCTCAGGC CAGGCACCAC AGCTCTTCCA   1560

30 CTCACAGACT CTCACAACTG CACCCTTGCC GGGCACCACT CCACTGTATC CCTCCCCCAT   1620

   GACTCCCATG ACCCCCATCA CTCCTGCCAC GCCAGCTTCG GAGAGTTCTG GGATTGTACC   1680

35 GCAGCTGCAA AATATTGTAT CCACAGTGAA TCTTGGTTGT AAACTTGACC TAAAGACCAT   1740

   TGCACTTCGT GCCCGAAACG CCGAATATAA TCCCAAGCGG TTTGCTGCGG TAATCATGAG   1800

   GATAAGAGAG CCACGAACCA CGGCACTGAT TTTCAGTTCT GGGAAAATGG TGTGCACAGG   1860

40 AGCCAAGAGT GAAGAACAGT CCAGACTGGC AGCAAGAAAA TATGCTAGAG TTGTACAGAA   1920

   GTTGGGTTTT CCAGCTAAGT TCTTGGACTT CAAGATTCAG AACATGGTGG GGAGCTGTGA   1980

45 TGTGAAGTTT CCTATAAGGT TAGAAGGCCT TGTGCTCACC CACCAACAAT TTAGTAGTTA   2040

   TGAGCCAGAG TTATTTCCTG GTTTAATCTA CAGAATGATC AAACCCAGAA TTGTTCTCCT   2100

   TATTTTTGTT TCTGGAAAAG TTGTATTAAC AGGTGCTAAA GTCAGAGCAG AAATTTATGA   2160

50 AGCATTTGAA AACATCTACC CTATTCTAAA GGGATTCAGG AAGACGACGT AATGGCTCTC   2220

   ATGTACCCTT GCCTCCCCCA CCCCCTTCTT TTTTTTTTTT TAAACAAATC AGTTTGTTTT   2280

55 GGTACCTTTA AATGGTGGTG TTGTGAGAAG ATGGATGTTG AGTTGCAGGG TGTGGCACCA   2340
```

52

GGTGATGCCC TTCTGTAAGT GCCCCTTCCG GCATCCCGGA ATTCCTGCAG CCCAACGCGG   2400

CCGCTTCGAG GGATCTTTGT GAAGGAACCT TACTTCTGTG GTGTGACATA ATTGGACAAA   2460

CTACCTACAG AGATTTAAAG CTCTAAGGTA AATATAAAAT TTTTAAGTGT ATAATGTGTT   2520

AAACTACTGA TTCTAATTGT TTGTGTATTT TAGATTCCAA CCTATGGAAC TGATGAATGG   2580

GAGCAGTGGT GGAATGCCTT TAATGAGGAA AACCTGTTTT GCTCAGAAGA AATGCCATCT   2640

AGTGATGATG AGGCTACTGC TGACTCTCAA CATTCTACTC CTCCAAAAAA GAAGAGAAAG   2700

GTAGAAGACC CCAAGGACTT TCCTTCAGAA TTGCTAAGTT TTTTGAGTCA TGCTGTGTTT   2760

AGTAATAGAA CTCTTGCTTG CTTTGCTATT TACACCACAA AGGAAAAAGC TGCACTGCTA   2820

TACAAGAAAA TTATGGAAAA ATATTCTGTA ACCTTTATAA GTAGGCATAA CAGTTATAAT   2880

CATAACATAC TGTTTTTTCT TACTCCACAC AGGCATAGAG TGTCTGCTAT TAATAACTAT   2940

GCTCAAAAAT TGTGTACCTT TAGCTTTTTA ATTTGTAAAG GGGTTAATAA GGAATATTTG   3000

ATGTATAGTG CCTTGACTAG AGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT   3060

TGCTTTAAAA AACCTCCCAC ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT   3120

TGTTGTTAAC TTGTTTATTG CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA   3180

TTTCACAAAT AAAGCATTTT TTTCACTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA   3240

TGTATCTTAT CATGTCTGGA TCC   3263

(2) INFORMATION FOR SEQ ID NO: 16:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 371 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (ix) FEATURE:
        (A) NAME/KEY: Protein
        (B) LOCATION: 1..371

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

    Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Val Met Gly Ser Ser
    1   5   10   15

    His His His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser His
                20   25   30

53

Met Asp Gln Asn Asn Ser Leu Pro Pro Tyr Ala Gln Gly Leu Ala Ser
        35   40   45

Pro Gln Gly Ala Met Thr Pro Gly Ile Pro Ile Phe Ser Pro Met Met
        50   55   60

Pro Tyr Gly Thr Gly Leu Thr Pro Gln Pro Ile Gln Asn Thr Asn Ser
65   70   75   80

Leu Ser Ile Leu Glu Glu Gln Gln Arg Gln Gln Gln Gln Gln Gln Gln
                85   90   95

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
        100   105   110

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Ala
        115   120   125

Val Ala Ala Ala Ala Val Gln Gln Ser Thr Ser Gln Gln Ala Thr Gln
        130   135   140

Gly Thr Ser Gly Gln Ala Pro Gln Leu Phe His Ser Gln Thr Leu Thr
145   150   155   160

Thr Ala Pro Leu Pro Gly Thr Thr Pro Leu Tyr Pro Ser Pro Met Thr
                165   170   175

Pro Met Thr Pro Ile Thr Pro Ala Thr Pro Ala Ser Glu Ser Ser Gly
                180   185   190

Ile Val Pro Gln Leu Gln Asn Ile Val Ser Thr Val Asn Leu Gly Cys
                195   200   205

Lys Leu Asp Leu Lys Thr Ile Ala Leu Arg Ala Arg Asn Ala Glu Tyr
        210   215   220

Asn Pro Lys Arg Phe Ala Ala Val Ile Met Arg Ile Arg Glu Pro Arg
225   230   235   240

Thr Thr Ala Leu Ile Phe Ser Ser Gly Lys Met Val Cys Thr Gly Ala
                245   250   255

Lys Ser Glu Glu Gln Ser Arg Leu Ala Ala Arg Lys Tyr Ala Arg Val
        260   265   270

Val Gln Lys Leu Gly Phe Pro Ala Lys Phe Leu Asp Phe Lys Ile Gln
        275   280   285

Asn Met Val Gly Ser Cys Asp Val Lys Phe Pro Ile Arg Leu Glu Gly
        290   295   300

Leu Val Leu Thr His Gln Gln Phe Ser Ser Tyr Glu Pro Glu Leu Phe
305   310   315   320

Pro Gly Leu Ile Tyr Arg Met Ile Lys Pro Arg Ile Val Leu Leu Ile

54

	325	330	335

Phe Val Ser Gly Lys Val Val Leu Thr Gly Ala Lys Val Arg Ala Glu

	340	345	350

5

Ile Tyr Glu Ala Phe Glu Asn Ile Tyr Pro Ile Leu Lys Gly Phe Arg

	355	360	365

Lys Thr Thr

10   370

(2) INFORMATION FOR SEQ ID NO: 17:

15   (i) SEQUENCE CHARACTERISTICS:
           (A) LENGTH: 18 amino acids
           (B) TYPE: amino acid
           (C) STRANDEDNESS: single
           (D) TOPOLOGY: linear

20

    (ii) MOLECULE TYPE: protein

    (ix) FEATURE:
25   (A) NAME/KEY: Protein
           (B) LOCATION: 1..18

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

30

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro
1   5   10   15

Arg Gly

55